Reproducibility in Science
Improving the Standard for Basic and Preclinical Research

C. Glenn Begley, John P.A. Ioannidis

Abstract: **Medical and scientific advances are predicated on new knowledge that is robust and reliable and that serves as a solid foundation on which further advances can be built. In biomedical research, we are in the midst of a revolution with the generation of new data and scientific publications at a previously unprecedented rate. However, unfortunately, there is compelling evidence that the majority of these discoveries will not stand the test of time. To a large extent, this reproducibility crisis in basic and preclinical research may be as a result of failure to adhere to good scientific practice and the desperation to publish or perish. This is a multifaceted, multistakeholder problem. No single party is solely responsible, and no single solution will suffice. Here we review the reproducibility problems in basic and preclinical biomedical research, highlight some of the complexities, and discuss potential solutions that may help improve research quality and reproducibility. (*Circ Res*. 2015;116:116-126. DOI: 10.1161/CIRCRESAHA.114.303819.)**

Key Words: funding ■ journals ■ research integrity ■ universities

Problem

As physicians and scientists, we want to make a contribution that alters the course of human health. We all want to make our mark by discovering something that really makes a difference. Yet most of the common diseases that we study are exceedingly complex. Our knowledge is fragmentary. Many, perhaps most of our models are naive and our constructs are often rough approximations of the truth. Eventually, we may see in our own research efforts and results what we want to see: promising findings and nice discoveries. This seeming success may not be reproducible. Yet robust, reproducible research is the foundation on which advances are built.

Against the reality of the diseases we want to cure, there is a real opportunity that we can more readily address. Although this would involve widespread changes and demand a critical re-evaluation of our processes, this is an opportunity that could have substantial benefit. The opportunity is to introduce, demand, and reward a level of rigor and robustness in designing, conducting, reporting, interpreting, validating, and disseminating research that is currently lacking from many areas of biomedical research.

Over the recent years, there has been an increasing recognition of the weaknesses that pervade our current system of basic and preclinical research. This has been highlighted empirically in preclinical research by the inability to replicate the majority of findings presented in high-profile journals.[1–3] The estimates for irreproducibility based on these empirical observations range from 75% to 90%. These estimates fit remarkably well with estimates of 85% for the proportion of biomedical research that is wasted at-large.[4–9] This irreproducibility is not unique to preclinical studies. It is seen across the spectrum of biomedical research. For example, similar concerns have been expressed for observational research where zero of 52 predictions from observational studies were confirmed in randomized clinical trials.[10–12] At the heart of this irreproducibility lie some common, fundamental flaws in the currently adopted research practices. Although disappointing, this experience should probably not be surprising, and it is what one would expect also theoretically for many biomedical research fields based on how research efforts are conducted.[13]

Basic and preclinical research is particularly important because it forms the foundation on which future studies are built. It is preclinical research that provides the exciting, new ideas that will eventually find their way into clinical studies and new drugs that provide benefit to humankind. Yet this preclinical research is poorly predictive of ultimate success in the clinic.[14,15] And it is observational research that both attracts immediate public attention and often provides the hypothesis on which interventional studies are based.

Given the controversy that has surrounded this issue, it is important to note at the outset that these concerns about the irreproducibility of science do not invalidate the validity or legitimacy of the scientific method. Rather it is the rigorous, careful application of the scientific method that has translated into genuine improvements in human health and provided the substantial benefits we have enjoyed over recent decades: the

Original received July 27, 2014; revision received August 22, 2014; accepted September 15, 2014. In November, 2014, the average time from submission to first decision for all original research papers submitted to *Circulation Research* was 13.96 days.

From the TetraLogic Pharmaceuticals Corporation, Malvern, PA; and Departments of Medicine and Health Research and Policy, Stanford University School of Medicine, Department of Statistics, Stanford University School of Humanities and Sciences, and Meta-Research Innovation Center at Stanford (METRICS), CA.

Correspondence to John P.A. Ioannidis, Stanford University, MSOB X306, Stanford, CA 94305. E-mail jioannid@stanford.edu

Nonstandard Abbreviations and Acronyms
None.

investment in research has yielded dramatic improvement in outcomes in infectious diseases, cardiovascular disease, oncology, rheumatic diseases, and many other conditions.

At the outset, it is also important to acknowledge that not all biomedical research is assumed or expected to result in findings that have application for human health. Excellent ideas and excellent research can lead to a dead-end. That is simply the nature of research. Efficiency of 100% and waste of 0% is unlikely to be achievable. However, there is probably substantial room for improvement.

Finally, the fact that this self-critical debate is occurring openly within the scientific community reinforces the strength of our system.

Why Is This Important?

The key challenge is to ensure the most efficient, effective use of precious research funds. In the preclinical arena, it has become increasingly clear that the majority of preclinical research is unable to be reproduced, including by the original authors themselves.

This has long been known to be the case by large biotechnology and pharmaceutical companies that routinely seek to reproduce a scientific claim before commencing a research project. The issue is more problematic for the smaller companies and those who provide funding for these earlier stage companies, where, in proportion to their resources, the cost of attempting to reproduce a study can be substantial both in terms of money and time.[16] Early, preclinical studies are important because they potentially form the foundation on which later studies are built to evaluate specific drugs, and other interventions, and they can also provide the framework for biomarker analyses that can be used to focus treatments more precisely.

What is difficult to quantify is the opportunity cost associated with studies that fail to replicate. Investigators who pursue a provocative and exciting idea but that is not based on robust data will consume wasted hours for an idea that is ultimately discarded. For ideas that proceed into late stages of clinical evaluation or are even adopted in clinical practice only to be discarded subsequently,[17] the cost can be enormous.

Addressing this concern is central to ensuring ongoing confidence of and support from the public. It is key that the public, who directly and indirectly provide the money to fund our research efforts and who are our most important ambassadors in advocating for the importance of scientific investigation, are confident of the processes we have in place. They need to know there is real value in the data that ultimately emerges. They, after all, are the ultimate recipients reaping the benefits of the biomedical research enterprise.

What Constitutes Reproducibility?

There is no clear consensus at to what constitutes a reproducible study. The inherent variability in biological systems means there is no expectation that results will necessarily be precisely replicated. So it is not reasonable to expect that each component of a research report will be replicated in perfect detail. However, it seems completely reasonable that the one or two big ideas or major conclusions that emerge from a scientific report should be validated and with-stand close interrogation.

What has shaken many in the field is not that investigators are unable to precisely reproduce an experiment. That is to be expected. What is shocking is that in many cases, the big idea or major conclusion was not confirmed simply when experiments were performed by the same investigators when blinded to their test samples versus control samples.[2] The explanation for this was evident when the precise methodology of the experiments was reviewed. Investigators typically performed their experiments in a nonblinded fashion, so they were able to see what they were anticipating to see, and their research bias was thus able to be confirmed.[18] Observer bias has long been recognized to be a problem in preclinical studies and beyond, so this result should not be surprising.[19] Confirmation bias in scientific investigation unavoidably makes even the best scientists prone to try to find results or interpretations that fit their preconceived ideas and theories.[20,21]

In addition, empirical assessments of preclinical studies showed an array of problems, including, but not limited to, the fact that there was a failure to repeat experiments, to use legitimate controls, to validate reagents, and use appropriate statistical tests. On top of that, investigators often selected the best experiment rather than referencing the entire data set. These practices conspired to ensure that not only could individual experiments not be replicated, but that the main conclusion of the article was not substantiated.[18]

Furthermore, it was notable that several nonreproducible publications had been cited many hundreds of times. Clinical studies were initiated based on that work.[2] But the authors of the secondary publications had not sought to actually reproduce or falsify the findings of the original papers.

There is an acknowledged tension here. It is reasonable to expect that there will be some level of uncertainty and irreproducibility as investigators genuinely push the boundaries of current knowledge and extend into the unknown. But it is also reasonable to expect that standard research procedures, such as experimenter-blinding, will be used and that irreproducibility will be seen in the minority rather than the majority of cases.

Has Much Changed?

The inability to reproduce research findings is probably a long-standing problem. In the past, investigators would share privately the work that others could not reproduce. This was a key part of scientific exchange that typically took place in the corridors or bars at scientific meetings.

This is an issue across the breadth of biomedical and social sciences (for illustrative examples, see Tables 1 and 2). However, the burgeoning number of high-profile journals has given voice to many papers that, in the past, would not have received the same level of profile and exposure. Now the sheer magnitude of high-profile studies creates a challenge for any

Table 1. Examples of Some Reported Reproducibility Concerns in Preclinical Studies

Author	Field	Reported Concerns
Ioannidis et al (2009)[22]	Microarray data	16/18 studies unable to be reproduced in principle from raw data
Baggerly et al (2009)[23]	Microarray data	Multiple; insufficient data/poor documentation
Sena et al (2010)[24]	Stroke animal studies	Overt publication bias: only 2% of the studies were negative
Prinz (2011)[1]	General biology	75% to 80% of 67 studies were not reproduced
Begley & Ellis (2012)[2]	Oncology	90% of 53 studies were not reproduced
Nekrutenko & Taylor(2012)[25]	NGS data access	26/50 no access to primary data sets/software
Perrin (2014)[26]	Mouse, in-vivo	0/100 reported treatments repeated positive in studies of ALS
Tsilidis et al (2013)[27]	Neurological studies	Too many significant results, overt selective reporting bias
Lazic & Essioux (2013)[28]	Mouse VPA model	Only 3/34 used correct experimental measure
Haibe-Kains et al (2013)[29]	Genomics/cell line analysis	Direct comparison of 15 drugs and 471 cell lines from 2 groups revealed little/no concordant data
Witwer (2013)[30]	Microarray data	93/127 articles were not MIAME compliant
Elliott et al (2006)[31]	Commercial antibodies	Commercial antibodies detect wrong antigens
Prassas et al (2013)[32]	Commercial ELISA	ELISA Kit identified wrong antigen
Stodden et al (2013)[33]	Journals	Computational biology: 105/170 journals noncompliant with National Academies recommendations
Baker et al (2014)[34]	Journals	Top tier fail to comply with agreed standards for animal studies
Vaux (2012)[35]	Journals	Failure to comply with their own statistical guidelines

ALS indicates amyotrophic lateral sclerosis; MIAME, minimum information about a microarray experiment; NGS, next generation sequencing; and VPA, valproic acid (model of autism).

investigator to remain current. The number of publishing scientists has grown over the years, with over 15 million scientists publishing ≥1 article that was indexed in Scopus in the period 1996–2011.[49] Biomedical research is the most prolific scientific field in this regard. It is practically impossible for even the most knowledgeable expert to maintain direct knowledge of the work done by so many other scientists, even when it comes to his/her core discipline of interest.

This issue is also amplified by the public appetite for new advances. Over recent years, an increasingly informed, vocal and vigilant patient–advocate community has emerged. These groups have played an important role in helping further raise the profile of medical research. They have effectively argued for ongoing research investment. At the same time, they have maintained a clear focus on research productivity and expect to see breakthrough discoveries effectively translated into improved medicines. And a hungry media is happy to trumpet these breakthroughs to a waiting public.

Why Is This Happening?

The vast majority of academic investigators, industry researchers, journal editors, investors, government, and certainly patients want and expect to see valuable research dollars translated into new therapies. The expectation is well founded: the evidence that biomedical research has had a profound effect on human health is undeniable. To be clear, the problem under discussion here is not one of scientific fraud. It is not a failure of the scientific method. It is a consequence of a system that is willing to overlook and ignore lack of scientific rigor and instead reward flashy results that generate scientific buzz or excitement. It is mostly an issue of how priorities are balanced and a failure to rigorously apply standard scientific methods in an increasingly competitive research environment when scientists are scrambling to get their share of a dwindling national research budget.[50] Although fraud is rare, use of questionable research practices seems to affect the majority of researchers.[51]

Table 2. Additional Basic Science Fields Where Concerns Regarding Reproducibility Have Been Raised

Discipline	Issues Raised	Author
Neuroscience	Low statistical power; small sample size	Button et al (2013)[36]
Pharmacology	Lack of training, lack of statistical power, blinding, hypothesis, requisite PK studies, randomization, dose-response, controls, prospective plan, validation, independent replication, and selection of doses that are not tolerable in humans	Henderson et al(2013)[37]; Kenakin et al (2014)[38]; McGonigie et al (2014)[39]; Winquist et al (2014)[40]; Marino (2014)[41]
Genomics/bioinformatics	Irreproducibility of high-profile studies	Sugden et al (2013)[42]
Stem cell biology	Lack of reliable, quality data	Plant & Parker (2013)[43]
Oncology, in vitro testing	Use of clinically unachievable concentrations	Smith & Houghton(2013)[44]
Chemistry lead-discovery	Artifacts; false positives and negatives	Davis & Erlanson (2013)[45]
Computational biology	10 common errors	Sandve et al (2013)[46]
Pathology/Biomarkers	Biospecimen quality	Simeon-Dubach et al (2012)[47]
Organizational psychology	Suppression of negative studies	Kepes & McDonald (2103)[48]
Observational research	0/52 hypotheses confirmed in randomized Trials	Young & Karr (2011)[11]

In the preclinical arena, there seems to be a wide-spread conscious or unconscious belief that a rigorous research process, that follows what most would consider standard scientific methodology (blinding, repeating experiments, inclusion of positive and negative controls, use of validated reagents, etc.), may stifle the creative, innovative act of discovery. That is clearly not the case. An unexpected, unexplained observation that serves as the first hint of something new should be tested rigorously. It should be repeated and carefully confirmed before it is announced to the world.[52] Unfortunately, currently that is seldom the case. Instead, it would seem that many investigators feel the need to immediately rush into print with an unconfirmed, and unconfirmable, finding.

It is probably relevant to note the evidence that investigators who are assured of more consistent, stable laboratory funding are less likely to succumb to the pressures of fraud.[53,54] It seems reasonable to extrapolate and suggest that may also be the case for investigators who enjoy similar guaranteed research funding, they too may be less inclined to take shortcuts with their experiments and use questionable or substandard research practices. Perhaps somewhat paradoxically then, it is reasonable to expect that if research funding per capita (per investigator) decreases further, we should expect that this problem will be even more evident: the need to be first, with its concomitant rewards, will trump the need to be right.

What Can be Done?

Addressing this challenge may require a multipronged approach, there is no single solution, and there are several initiatives underway that may have a positive effect. What seems to emerge repeatedly is a failure for individual researchers, reviewers, and editors to comply with agreed, well-established guidelines for the conduct of experimental research. While there is no direct incentive for investigators to comply, it seems that generally they do not.

Agreed Recommendations for Pharmacology Studies Are Overlooked

There have been calls for a more rigorous approach to experimental design with some disciplines devoting considerable effort to addressing the inability to translate preclinical research into clinical success. For example, in the field of pharmacology in an effort to improve the quality of animal studies, investigators have argued for a focus on hypothesis-testing research,[40] a prospective, rigorous, research plan for preclinical studies,[38] training and more appropriate use of statistics,[41] and performance of requisite studies, such as pharmacokinetic analyses before any evaluation of potential efficacy.[39] These recommendations are widely endorsed and supported. Henderson et al[37] came to similar conclusions after their review of the literature. Among the most common recommendations included blinding of outcome assessment, randomized allocation of animals, power calculation to determine sample size, use of positive and negative controls, determination of dose-response, replication in different models, and independent replication.

Although these are key to the appropriate interpretation of animal studies, they are still widely neglected. One particularly troubling study[28] found that only 9% of studies used the correct experimental unit in their analyses and, further, that the behavioral variation between litters of mice was greater than the reported treatment effect. This failure to adopt well-recognized, widely supported guidelines begs the question: how does the community ensure that key recommendations are integrated into the research effort? How is this monitored?

Data Disclosure and Omics Research: Guidelines Are Ignored

Another area that has received attention is the analysis of large data sets and the computing tools that are required to analyze them. There are several common issues with omics research.

It was recognized over a decade ago that experiments involving array technology were potentially fraught with problems. To address this, a requirement for mandatory full data deposition was recommended in 2001 and quickly adopted by many journals. This minimum information about a microarray experiment standard has been widely accepted by investigators and journals. However, despite being accepted as the desired norm over a decade ago, compliance remains a problem. An analysis of data from high-profile microarray publications that appeared between 2005 and 2006 in Nature Genetics reproduced the results of only 2 of 18 reanalyzed papers. The principal reason for failure was lack of availability of original raw data.[22] A study of 127 articles on microarray studies published between July 2011 and April 2012 revealed that ≈75% were still not minimum information about a microarray experiment compliant. Furthermore, reanalysis of data often did not support the original conclusions.[30] Similar observations have been made by others.[23,25] An analysis of 500 papers in the 50 top journals across scientific fields (those with highest impact factors) revealed only 9% deposited full primary data online.[55]

It is noteworthy that both the National Science Foundation and the National Institutes of Health have strongly worded statements requiring data set disclosure and software sharing, but neither is consistently enforced: compliance still remains at the discretion of the investigators.

In addition, the majority of journals surveyed (105 of 170 journals) in a recent evaluation failed to comply with the 2003 National Academies report requiring that journals clearly and prominently state (in the instructions for authors and on their Web sites) their policies for distribution of publication-related materials, data, and other information.[33] Similar calls have come from many directions and with a plea to support only those journals that support reproducible research.[56–62] These studies reinforce the importance of access to primary data. Moreover, the data suggests that a reluctance to provide such access is a marker of a lower quality study.

There are additional issues with omics research beyond access to the data and specific recommendations have been proposed to address these, in particular regarding the need to perform independent validations of all findings that emerge from these analyses.[63]

Substantive issues still persist in studies involving large data sets. As a recent example, when the activity of 15 oncology drugs was directly compared in large pharmacogenomics studies from 2 independent data sets, one generated at the Dana-Faber Cancer Institute and the other at the Massachusetts General Hospital, there was little or no

consistency, even though the same 471 cell lines were examined.[29,64] Although the lack of consistency might have been attributed to differences in the cell lines that were used, this seems unlikely. Rather, it probably represents a lack of standardization of the experimental assays and methods of analysis.[29]

We Get What We Incentivize

It is interesting to observe what has happened in the psychology literature. Some of the most provocative psychological experiments were not able to be replicated. For example, the alleged effect of positive or negative priming was not confirmed when experiments were rigorously repeated.[65] The publication of this negative result is unusual. Instead, investigators have noted that the rate of positive results in psychological science (as in many biomedical fields) is approaching 90% to 100%.[66] Most hypotheses are confirmed, and skeptics have concluded that we are either "approaching omniscience or our journals are publishing an unrepresentative sample of completed research."[48] It is likely that this bias toward studies that confirm a favorite hypothesis spans the scientific spectrum.[2,67–69] An analysis of 525 preclinical stroke studies revealed that only 2% reported a lack of effect on stroke, which led the authors to conclude that serious publication bias may exist.[24] In another analysis of over 4445 data sets involving animal studies of neurological diseases, it was similarly concluded that perhaps the majority of the data were either suppressed or recast in a way that truly negative studies would be published as positive results—there were just too many positive results published to be true.[27] Correspondingly, although the large majority of interventions for neurological diseases seemed effective in animal studies, few had favorable results when tested in humans.

Further, the recognition that in psychological research using several smaller, underpowered samples is more likely to provide a positive result through selective analysis and outcome reporting than using a larger appropriately powered sample[70] also permeates other biomedical fields.[71] Similarly, the powerful incentives that favor novelty over replication are certainly not unique to psychological studies.[72] In an attempt to proactively address these concerns, the reproducibility project has been created. This is a large-scale, cooperative effort to examine reproducibility in psychology studies[73] and could serve as a template for other disciplines.

Several thoughtful suggestions have been made by many investigators, some of which are listed in Table 3. Although a systematic review of the literature was not performed, there is substantial concordance in terms of these recommendations, with a commonality across diverse disciplines and research stage. The fundamental problem with most, if not all, of these proposals is the requirement for investigators, institutions, and journals to willingly comply: it is not at all clear how reasonable recommendations will be implemented or monitored while they remain voluntary. Conversely, were they to be mandatory, then one has to examine carefully how they would be enforced and by whom. The details of how to make these changes work can have a major effect on their efficiency.

Additional Initiatives

There are several other initiatives underway to try and aid this process. We list here a few examples. The Neuroscience Information Framework was created to define resources available to the research community to further neuroscience research.[79] Investigators studying spinal cord injury have defined the minimal information about a spinal cord injury experiment in an attempt to improve the outcomes of spinal cord injury research.[80] They have proposed the inclusion of specific experimental details in addition to more general elements of study design, such as experimenter blinding, randomization of animals, cohorts of sufficient size, inclusion of controls.[18,81]

The Global Biological Sciences Institute aims to establish a set of harmonized, consensus-based standards that can be applied to commonly used research tools. An agreed set of validated tools would represent a significant advance. But even with validated tools, having them accepted will be a challenge. This is illustrated by the MDA-MB-435 cell line that has hundreds of citations falsely identifying it as a breast cell. Despite definitive reports identifying this as a melanoma cell,[82,83] at the time of writing, since 2010, there have been over 170 reports continuing to falsely identify it as a breast cell and only 47 reports correctly identifying it as a melanoma cell (Begley unpublished). A curated data set that is publicly available with a list of readily searchable reagents could go a long way to helping address this type of problem.

Another recent project is the Reproducibility Initiative, which seeks to examine the replicability of the top 50 most impactful cancer biology studies published between 2010 and 2012.[84] This initiative is exploratory and may offer useful insights. It potentially offers the opportunity to debate and clarify within the scientific community what constitutes adequate and appropriate replication, who should perform these replication studies, and how they are best performed.[52]

Academia.edu is taking a different approach, attempting to build an alternative publication system where papers are made available promptly and with the peer review process taking place postpublication. They also hope to make research freely available.

These efforts are all still at an embryonic stage, and although their value remains to be determined, it is heartening to see the broad-based approach that is being adopted by the scientific community.

The National Institute of Health has indicated that it too will take several steps, including additional training within the National Institute of Health intramural program, a more systematic review of grant applications, more transparent access to data, an online forum for discussion of published papers along with other measures under consideration.[50] Of these, perhaps the most important is the focus on education. This is a crucial initiative. It is essential that the next generation of researchers is better trained in terms of design, execution, and interpretation of experiments. A key component of this effort will need to be instruction in terms of appropriate selection and cleaning of data for presentation. This initiative will help address the concern that junior researchers may not receive adequate personalized training or mentoring if they are

Table 3. Some Proposals to Improve Experimental Rigor and Quality in Preclinical Research

Proposal	Author
Editors solicit replication bids	Wagenmakers and Forstman (2014)[74]
Plea to improve editorial standards	Multiple, eg, Kraus (2014),[75] and Refs. 56-72
Reward quality rather than quantity	Kraus (2014)[75]
Emphasis on hypothesis testing research	Winquist et al (2014)[40]
Prospective, rigorous experimental plan	Kenakin et al (2014)[38]
Improved understanding of statistics	Marino (2014)[41]; Vaux (2012)[35]
Improved experimental design	Henderson et al (2013)[37]
Systematic reviews of animal studies	Hooijmans & Ritskes-Hoitinga (2013)[76]
Use clinically relevant concentrations	Smith & Houghton (2013)[44]
Consider litter effects	Lazic & Essioux (2013)[28]
Recommendations to improve computational biology	Sandve et al (2013)[46]
Focus on reproducibility in training, grants, journals	LeVeque et al (2012)[61]
Pathology: Biospecimen quality control	Simeon-Dubach et al (2012)[47]
Microarray analyses: Provide data access	Witwer (2013)[30]
Psychology: open data, methods and workflow	Nosek et al (2012)[72]
Meta-analyses of animal data	Macleod et al (2004)[77]
Judge academics on quality, reproducibility, sharing	Ioannidis et al (2014)[6]
Greater institutional responsibility	Chan et al (2014)[9]
Apply greater skepticism to new technologies	Glaeser (2006)[78]

in laboratories where the principal investigator is unable to provide individual attention on a regular basis.

Realistic Prospect for Change?

Despite the clear challenges that we face, it is worth recalling that substantive changes have taken place over recent decades in clinical trials processes, and similar progress may also be achieved for preclinical research. Although it has taken a long time to reach wider consensus about the importance of proper randomization, blinding (when feasible and appropriate), and registration in clinical trials, these are now widely accepted as the norm, and failure to comply (eg, nonregistration) is penalized (eg, by inability to publish in most major journals). When these changes were being introduced, some physicians expressed concern, at least privately that these changes would stifle clinical research. In fact, this has introduced a new level of reliability and confidence in the clinical research effort. There are still multiple challenges of course. The public availability of clinical trial data remains suboptimal.[85,86] Research protocols and full study reports remain difficult to access. Transparent access to key data that underpins clinical conclusions remains to be broadly addressed.[9] Plus many clinical trials remain nonrandomized without any good reason, especially in phase I and II research.[87] However, despite these real concerns, the current status of planning, execution, and interpretation of clinical trials is different and probably substantially improved to what it was several decades ago and vastly superior to the situation in preclinical research.

At this time, the level of control, rigor, and accountability seen in clinical trials (even with their residual deficiencies) is currently unimaginable in the preclinical arena[88] where much still remains at the discretion of the individual investigator. Although it seems unlikely that this pattern can continue, the notion of increased oversight of preclinical studies has long been proposed.[89,90] The current model of investigator self-regulation and self-censoring does not seem to be serving the scientific community well enough. Although clinical and preclinical research do have major differences, many of the lessons learnt from clinical research may also be considered for more systematic application in preclinical research.

In addition to the changes that have taken place in clinical practice, it is important to recall the changes that have been successfully introduced into basic and preclinical research. Many of these are now integrated into standard research practice and are accepted as routine procedures. As a result, there are now appropriate limits on handling of radioactive materials, recommendations regarding DNA cloning, committees to ensure appropriate research access to human tissues, safeguards to protect investigators when using human samples, guidance regarding use of retroviruses, regulations on experiments with embryonic stem cells, and committees to ensure appropriate treatment of animals in research. The introduction of each of these was associated with some increase in bureaucracy and significant reluctance on the part of some investigators: the introduction of regulations and controls for DNA cloning was hotly debated in the early 1970s. At that time, Watson strongly opposed regulatory protocols,[91] whereas other scientists, including Singer and Berg, took the lead on promoting a responsible approach to the young field.[92,93] There are obvious parallels in the current debate regarding the extent and importance of reproducible science.[64]

Once guidelines and regulations are introduced, these measures have typically remained in place. However, at least for some (eg, controls on DNA cloning and embryonic stem cells), controls have become more relaxed in the light of new scientific information and a changing social and biomedical environment.[92–95]

Although it is not possible to demonstrate a direct improvement in the research enterprise as a result of these changes, for those who recall the time before their introduction, it seems self-evident that there is improved investigator safety as a result of training in use of radioactive materials and safeguards regarding potential infectious agents. It seems intuitively obvious that these changes have resulted in improved animal care and better protection for patients' rights. But there is no doubt that this came at a cost in terms of increased bureaucracy, increased demands on investigators' time, and increased institutional responsibility.

Role of the Editors, Reviewers, and Journals

It has been argued that editors, reviewers, and the journals must take substantial responsibility for the current situation.[96] Many of the perverse incentives that drive scientific

Table 4. Issues That Could Be Addressed by a Policy of Good Institutional Practice for Basic Research

Focus	Proposal
Students/post-doctoral fellows	Core training in experimental methods and experimental design; data selection; data analysis; blinding; inclusion of controls; statistical interpretation; reagent validation; experimental replicates and repeats
	Mentoring provided by senior colleague from independent department
Investigator	Requirement that subjective end points are assessed by blinded investigators
	Compulsory refresher courses on experimental design; data selection; inclusion of controls; data analysis; statistical interpretation; reagent validation; issues in emerging technologies
	Requirement to comply with Federal and Scientific community guidelines and recommendations
Institution	Guidelines for dealing with fraud
	Independent committee to review compliance
	Requirement that raw data will be made available on request
	Guidelines for recording of laboratory notebooks
	Random reviews of laboratory notebooks
	Transparent promotion process that weighs quality above flashy, nonreproducible research; rewards mentoring and training

publications, and thus build and sustain careers, could be more closely regulated at this level.[75]

Several journals have introduced new Guidelines for Authors and made specific proposals that attempt to address the problems in preclinical scientific methodology identified above.[62,97,98] These Guidelines attempt to increase the focus on replication, reagent validation, statistical methods, and so on. This should make it more difficult for papers to be accepted as the journals will require additional controls and validation. Complying with these requirements will increase the effort of investigators, but the literature may become more reliable. New standards should hopefully improve the quality of scientific publications. However, it is again worth noting the difference between having standards in place and ensuring those standards are met: this continues to prove to be a problem.[34,35]

Monitoring the effect of policy changes is warranted: there is no value in simply increasing the burden of bureaucracy. It is possible that some new policies may have the best intentions but may result in inadvertent adverse consequences. There is also the potential of normative responses where investigators focus their attention on satisfying whatever new checklist item is asked of them, but fail to improve the overall agenda of their research.

Given that most initiatives for making changes are done within single journals or scientific subfields, there is fragmentation and lack of consistency in the messages given and in the policies that are adopted. For example, despite the improvements in Guidelines to Authors, and despite the decades-long recognition of the need for investigator blinding,[19] there seems to still be a general reluctance to demand blinding of investigators even when subjective end points are assessed. As with this journal, even the journals that have reworked their Guidelines to Authors still do not demand blinding of investigators in the evaluation of subjective preclinical data. In contrast to clinical trials where blinding is not always feasible or practical, in most preclinical investigations, blinding should be straightforward to adopt (with few exceptions). This oversight is difficult to comprehend, but presumably it reflects a reluctance and resistance from the investigator community.

There have also been specific editorial proposals to focus on rewarding investigators for scientific quality rather than quantity, reward confirmatory experiments, and publishing negative data, perhaps even preferentially so.[99] The proposal to reconsider the need to find clinical relevance in early stage research is particularly intriguing.[75] Although these recommendations make intuitive sense, a major challenge is how they would be introduced and how their effect would be evaluated. Another suggestion is that editors could openly solicit replication studies for findings of particular relevance, with a bidding-process and requirement for rigorous experimental approach.[74] However, the laboratories that are likely to respond to such a request may not be the ideal laboratories to perform such studies.[52]

Although it is easy to point to the journal editors and reviewers as the principal area for improvement, this ignores the fact that the publication of data is the final step in a lengthy process. Reviewers can have no knowledge as to what data investigators have chosen to exclude. Reviewers cannot know whether data was strung together post hoc simply to create the best story. Of necessity, they take the work at face value. Further, there is little recognition for reviewers who diligently undertake their work: they have many demands on their time, and reviewing an article may be the lowest priority. In fact, the reward for conscientious reviewers may be a disproportionate increase in reviewing requests!

Finally, particularly in the top-tier journals, there is a clear emphasis on the exploratory investigations that are at the heart of this problem.[97] These are forgiven the same level of rigor that is demanded of hypothesis-testing studies,[97,98] yet in terms of their prognostic value, they are most directly comparable to Phase 1 clinical studies. Although the focus on exploratory investigations is understandable, after all they generate scientific buzz, they should be labeled for what they are and viewed with the high level of skepticism they deserve.[98]

Back to the Beginning: Improving the Grant Review Process

Although a major part of the problem relates to our current system of scientific publishing, this is just the final stage of the research process. What we really need perhaps is a much more stringent focus on the grant approval process. Research

Table 5. Some Potential Recommendations

Funding Agencies, Investigators, Institutions, Journals

- Routine application of good scientific method (blinding, controls, repeats, presentation of representative data, reagent validation, adequate powering, etc)
- Demand and monitor compliance with consensus-based, peer-endorsed guidelines (eg, recommendations for animal pharmacology; MIAME; neuroscience studies, etc)
- Demand and monitor compliance with National Science Foundation and the National Institutes of Health requirements regarding data access

Funding agencies

- Provide more longer-term funding (people rather than projects)
- Fund projects to evaluate effect of compliance interventions
- Support reagent validation projects (antibodies; small molecules; siRNA, etc)
- Provide courses on scientific method for training junior investigators
- Monitor and reward reproducible, robust rather than flashy studies

Institutions

- Monitor and reward investigator-compliance with peer-generated guidelines and funding-agency requirements
- Monitor and reward reproducible, robust rather than flashy studies
- Support studies to evaluate effect of compliance interventions
- Provide compulsory courses on scientific method for junior and senior investigators

Journals

- Label exploratory investigations for what they are (ie, the equivalent of a phase 1 clinical study)
- Give greater weight to hypothesis-testing studies (the equivalent of a phase 2,3 clinical study)
- Encourage prepublication (eg via arXiv.org)

MIAME indicates minimum information about a microarray experiment.

funding agencies have primary responsibility for reductions in waste that emanates from how precious research funds are invested.[5]

As with those who review scientific papers for publication, Grant Committee Members are busy people who are seldom recognized or rewarded for a detailed, careful review. There is evidence that the grants that are funded have more similarity in their themes to the interests of the study section members than unfunded grants.[100] Grant reviewers may be more likely to suffer from allegiance-bias favoring grants that align with their own beliefs and interests. There is also some evidence that current grant review is inefficient and is not receptive of out-of-the box innovation. Different approaches to allocation of research funds have been proposed.[101] Many of those would focus more on judging the merit of people rather than projects.

In judging the merit of people, it is important to identify rigorous metrics that reward quality and reproducibility rather than volume and quantity. A PQRST index has been proposed for evaluating investigators[102] from the initials of productivity, quality, reproducibility, sharing, and translational potential. Consideration of these dimensions may improve the current situation. However, assessment of quality may be one of the most difficult aspects of judging the work and track-record of an investigator. Such assessment may require more transparent availability of information about the study design and conduct of previous studies by an investigator. It also requires some consensus on what are the key quality features that are more important in each field (eg, whether experiments were performed blinded, with appropriate reagents, experiments repeated, whether data sets are publicly available, etc).[18] In this regard, it is again worth noting the relative concordance of ideas across fields as to what constitutes quality research (Table 3). Quality indicators that acquire power to inform funding decisions are expected to improve: the adoption of a quality score would provide a powerful incentive for scientists to improve their performance specifically in the dimensions captured by that score. Optimistically, if these indicators are selected wisely and they represent the quality of the work well enough, there is a real possibility that the quality of proposed and conducted research will truly improve. More pessimistically, investigators may just see this as another checklist they have to satisfy that has no real effect on improving their research.[6] Again, a monitoring component would be key to include alongside a change in process that attempts to recognize quality. The general issue is currently under consideration by the National Institute of Health leadership[50] and provides some optimism that quality indicators will factor in to the grant review process going forward.

One simple proposal is that Committees should expect that all subjective analyses will be performed by blinded investigators. There should also be a requirement that reagents are validated, that a minimum number of replicates will be evaluated, and that appropriate controls will be included.[18,37] These requirements should be explicitly stated and include a rigorous defined research plan.[38] Unfortunately, these aspects of a research proposal are frequently regarded as trivial and are taken for granted: it is assumed that if the idea is a good one, that rigorous and careful scientific method will be used.

Institutional Responsibility

The principal responsibility for research findings rests with the investigator and their host institution. Currently, science operates under the trust me model that is no longer considered appropriate in corporate life nor in government.[72] It has been argued that there is also a need for institutions to take a greater leadership role.[9]

One proposal is that for Institutions that receive federal research funding, there should be an expectation that they will comply with Good Institutional Practice (analogous to, eg, Good Manufacturing Practice or Good Laboratory Practice). This would ensure that Institutions commit to and are recognized and rewarded for ensuring a minimum research standard among their employees. A disadvantage is that this will necessarily increase the burden on investigators and Institutions. It is essential to pick the right targets or what items to include under Good Institutional Practice and to avoid simply worsening research bureaucracy. In addition to improving the overall quality of our research endeavor, many of these recommendations would/should have immediate benefit to the Institution itself both in terms of the quality of personnel that they attract and the ability to commercialize their research (see Table 4 for illustrative examples).

Institutions could have a stated policy regarding their expectation for research quality conducted within their laboratories. For example, this may include a requirement that subjective end points are only evaluated by investigators blinded to the experimental arms. When the issue of blinding of experiments is discussed with principal investigators, a common concern is that multiple investigators are required for blinding to be effective. This of course is the point. Within an individual laboratory, this may require that students and postdoctoral fellows work together on a project. Although sharing a project and sharing data are contrary to the prevailing ethos in many laboratories, working together on the same experiment provides a level of cross-checking, cross-validation, and objectivity that would be beneficial.[9]

The institution has a responsibility to staff, students, and postdoctoral fellows. Compulsory annual refresher training for all principal investigators in experimental design, use of controls, use of validated reagents, data selection, statistical tests, and so on should be the norm. There should be similar training in experimental methods for junior researchers: this is an area where there is room for improvement even among highly regarded institutions.[103]

Institutions may also contribute toward improving transparency standards, for example, by adopting policies, indicating that raw data should be made available on request.

Institutions may also serve the goal of improving research credibility and efficiency if they adopt appointment and promotion standards that, instead of relying on publication in top tier journals as a surrogate for quality, recognizes the importance of reproducible research findings rather than flashy, unsubstantiated reports.[6,7]

Although many investigators are clearly able to self-regulate, self-monitor, self-censor, many others are not. For those investigators, we suggest their institution could have processes to ensure that there is an additional level of control that can act as a safety net to ensure the most appropriate scientific behaviors.

Conclusions

It is impossible to endorse an approach that suggests that we proceed with an ongoing research investment that is producing results the majority of which cannot be substantiated and will not stand the test of time.

It is time to rethink methods and standardization of research practices. These subjects are not nearly as superficially exciting as the exploratory studies that often seize wider interest, but their effect can be far more pervasive across the research enterprise.

Clearly not all investigators agree. Many agree with the view that we cannot afford to put programs of discovery on hold while we do a rethink on methods and standardization.[64] We offer a different perspective. If this means (and it probably does) putting some alleged advances on hold, that is completely appropriate if these touted advances are not really reproducible and truly useful. If this means (and it probably does) increasing the investment in research training and research infrastructure at the short-term expense of more bench research, that is also appropriate if the investment will eventually improve the yield of whatever bench research is done.

Corrective actions may need to involve several of the stakeholders discussed above in combination (see Table 5 for some specific recommendations). For example, if funding agencies make some right steps but these are not adopted by institutions or journals, the benefit may not be reaped. We also need to recognize that although there are many good ideas on how to improve research practices, their exact implementation can be a challenge. Evidence, ideally from experimental studies, on the effect of changes in research practices would be important to obtain and the effect of proposed changes should be properly monitored.

Acknowledgments

We are delighted to acknowledge the many colleagues who have contributed to this work. Their example, discussion, and critique over many years has been invaluable and is gratefully acknowledged.

Sources of Funding

The Meta-Research Innovation Center at Stanford (METRICS) is supported by a grant from the Laura and John Arnold Foundation.

Disclosures

C.G.B. is an employee and stock-holder in TetraLogic Pharmaceuticals Corporation and stock-holder in Amgen Inc.

References

1. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10:712. doi: 10.1038/nrd3439-c1.
2. Begley CG, Ellis L. Drug development: raise standards for preclinical research. *Nature*. 2012;483:531–533. doi: 10.1038/483531a.
3. Peers IS, Ceuppens PR, Harbron C. In search of preclinical robustness. *Nat Rev Drug Discov*. 2012;11:733–734. doi: 10.1038/nrd3849.
4. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374:86–89. doi: 10.1016/S0140-6736(09)60329-9.
5. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, Howells DW, Ioannidis JP, Oliver S. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014;383:156–165. doi: 10.1016/S0140-6736(13)62229-1.
6. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383:166–175. doi: 10.1016/S0140-6736(13)62227-8.
7. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, Michie S, Moher D, Wager E. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383:267–276. doi: 10.1016/S0140-6736(13)62228-X.

8. Al-Shahi Salman R, Beller E, Kagan J, Hemminki E, Phillips RS, Savulescu J, Macleod M, Wisely J, Chalmers I. Increasing value and reducing waste in biomedical research regulation and management. *Lancet*. 2014;383:176–185. doi: 10.1016/S0140-6736(13)62297-7.
9. Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzsche PC, Krumholz HM, Ghersi D, van der Worp HB. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 2014;383:257–266. doi: 10.1016/S0140-6736(13)62296-5.
10. Young SS, Bang H, Oktay K. Cereal-induced gender selection? Most likely a multiple testing false positive. *Proc Biol Sci*. 2009;276:1211–1212; discussion 1213. doi: 10.1098/rspb.2008.1405.
11. Young SS, Karr A. Deming, data and observational studies: a process out of control and needing fixing. *Significance*. 2011:9:122–126.
12. Young SS, Miller HI. Are medical articles true on health, disease? Sadly, not as often as you might think. *Genetic Engineering and Biotechnology News*; 2014: 34.
13. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2:696–701. doi: 10.1371/journal.pmed.0020124
14. Hackam DG, Redelmeier DA. Translation of research evidence from animals to humans. *JAMA*. 2006;296:1731–1732. doi: 10.1001/jama.296.14.1731.
15. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, Macleod M, Mignini LE, Jayaram P, Khan KS. Comparison of treatment effects between animal experiments and clinical trials: systematic review. *BMJ*. 2007;334:197–200. doi: 10.1136/bmj.39048.407928.BE.
16. Booth B, 2011 http://lifescivc.com/2011/03/academic-bias-biotech-failures/
17. Prasad V, Cifu A, Ioannidis JAP. Reversals of established medical practices evidence to abandon ship. *J Am Med Assoc* 2012;307:37–38. doi: 10.1001/jama.2011.1960.
18. Begley CG. Six red flags for suspect work. *Nature*. 2013;497:433–434. doi: 10.1038/497433a.
19. Rosenthal R, Lawson R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. *J Psychiatr Res*. 1964;2:61–72.
20. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol*. 1998;2:175–220.
21. Mynatta CR, Doherty ME, Tweney RD. Confirmation bias in a simulated research environment: an experimental study of scientific inference. *Q J Exp Psychol*. 1977;29:85–95.
22. Ioannidis JP, Allison DB, Ball CA, et al. Repeatability of published microarray gene expression analyses. *Nat Genet*. 2009;41:149–155. doi: 10.1038/ng.295.
23. Baggerly KA, Coombes KR. Deriving chemosensitivity from cell lines: forensic bioinformatics and reproducible research in high-throughput biology. *Ann Appl Stat*. 2009;3:1309–1334.
24. Sena ES, van der Worp HB, Bath PM, Howells DW, Macleod MR. Publication bias in reports of animal stroke studies leads to major overstatement of efficacy. *PLoS Biol*. 2010;8:e1000344. doi: 10.1371/journal.pbio.1000344.
25. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat Rev Genet*. 2012;13:667–672. doi: 10.1038/nrg3305.
26. Perrin S. Make mouse studies work. *Nature*. 2014;507:423–425.
27. Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, Al-Shahi Salman R, Macleod MR, Ioannidis JP. Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol*. 2013;11:e1001609. doi: 10.1371/journal.pbio.1001609.
28. Lazic SE, Essioux L. Improving basic and translational science by accounting for litter-to-litter variation in animal models. *BMC Neurosci*. 2013;14:37. doi: 10.1186/1471-2202-14-37.
29. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ, Quackenbush J. Inconsistency in large pharmacogenomic studies. *Nature*. 2013;504:389–393. doi: 10.1038/nature12831.
30. Witwer KW. Data submission and quality in microarray-based microRNA profiling. *Clin Chem*. 2013;59:392–400. doi: 10.1373/clinchem.2012.193813.
31. Elliott S, Busse L, Bass MB, Lu H, Sarosi I, Sinclair AM, Spahr C, Um M, Van G, Begley CG. Anti-Epo receptor antibodies do not predict Epo receptor expression. *Blood*. 2006;107:1892–1895. doi: 10.1182/blood-2005-10-4066.
32. Prassas I, Brinc D, Farkona S, Leung F, Dimitromanolakis A, Chrystoja CC, Brand R, Kulasingam V, Blasutig IM, Diamandis EP. False biomarker discovery due to reactivity of a commercial ELISA for CUZD1 with cancer antigen CA125. *Clin Chem*. 2014;60:381–388. doi: 10.1373/clinchem.2013.215236.
33. Stodden V, Guo P, Ma Z. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PLoS One*. 2013;8:e67111. doi: 10.1371/journal.pone.0067111.
34. Baker D, Lidster K, Sottomayor A, Amor S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for preclinical animal studies. *PLoS Biol*. 2014;12:e1001756. doi: 10.1371/journal.pbio.1001756.
35. Vaux D. Know when your numbers are significant. *Nature*. 2012;492:180–181.
36. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14:365–376. doi: 10.1038/nrn3475.
37. Henderson VC, Kimmelman J, Fergusson D, Grimshaw JM, Hackam DG. Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. *PLoS Med*. 2013;10:e1001489. doi: 10.1371/journal.pmed.1001489.
38. Kenakin T, Bylund DB, Toews ML, Mullane K, Winquist RJ, Williams M. Replicated, replicable and relevant-target engagement and pharmacological experimentation in the 21st century. *Biochem Pharmacol*. 2014;87:64–77. doi: 10.1016/j.bcp.2013.10.024.
39. McGonigle P, Ruggeri B. Animal models of human disease: challenges in enabling translation. *Biochem Pharmacol*. 2014;87:162–171. doi: 10.1016/j.bcp.2013.08.006.
40. Winquist RJ, Mullane K, Williams M. The fall and rise of pharmacology–(re-)defining the discipline? *Biochem Pharmacol*. 2014;87:4–24. doi: 10.1016/j.bcp.2013.09.011.
41. Marino MJ. The use and misuse of statistical methodologies in pharmacology research. *Biochem Pharmacol*. 2014;87:78–92. doi: 10.1016/j.bcp.2013.05.017.
42. Sugden LA, Tackett MR, Savva YA, Thompson WA, Lawrence CE. Assessing the validity and reproducibility of genome-scale predictions. *Bioinformatics*. 2013;29:2844–2851. doi: 10.1093/bioinformatics/btt508.
43. Plant AL, Parker GC. Translating stem cell research from the bench to the clinic: a need for better quality data. *Stem Cells Dev*. 2013;22:2457–2458. doi: 10.1089/scd.2013.0188.
44. Smith MA, Houghton P. A proposal regarding reporting of in vitro testing results. *Clin Cancer Res*. 2013;19:2828–2833. doi: 10.1158/1078-0432.CCR-13-0043.
45. Davis BJ, Erlanson DA. Learning from our mistakes: the 'unknown knowns' in fragment screening. *Bioorg Med Chem Lett*. 2013;23:2844–2852. doi: 10.1016/j.bmcl.2013.03.028.
46. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. *PLoS Comput Biol*. 2013;9:e1003285. doi: 10.1371/journal.pcbi.1003285.
47. Simeon-Dubach D, Burt AD, Hall PA. Quality really matters: the need to improve specimen quality in biomedical research. *J Pathol*. 2012;228:431–433. doi: 10.1002/path.4117.
48. Kepes S, McDaniel MA. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology-Perspectives on Science and Practice* 2013;6:252–268.
49. Boyack KW, Klavans R, Sorensen AA, Ioannidis JP. A list of highly influential biomedical researchers, 1996-2011. *Eur J Clin Invest*. 2013;43:1339–1365. doi: 10.1111/eci.12171.
50. Collins FS, Tabak LA. NIH plans to enhance reproducibility. *Nature*. 2014;505:612–613.
51. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*. 2009;4:e5738. doi: 10.1371/journal.pone.0005738.
52. Bissell M. The risks of the replication drive. *Nature*. 2013;503:333–334.
53. Fang FC, Casadevall A. Reforming science: structural reforms. *Infect Immun*. 2012;80:897–901. doi: 10.1128/IAI.06184-11.
54. Casadevall A, Fang FC. Reforming science: methodological and cultural reforms. *Infect Immun*. 2012;80:891–896. doi: 10.1128/IAI.06183-11.
55. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JP. Public availability of published research data in high-impact journals. *PLoS One*. 2011;6:e24357. doi: 10.1371/journal.pone.0024357.
56. Peng RD, Dominici F, Zeger SL. Reproducible epidemiologic research. *Am J Epidemiol*. 2006;163:783–789. doi: 10.1093/aje/kwj093.
57. Peng RD. Reproducible research and Biostatistics. *Biostatistics*. 2009;10:405–408. doi: 10.1093/biostatistics/kxp014.
58. Hyndman RJ. Encouraging replication and reproducible research. *Int J Forecasting*. 2010;26:2–3.
59. Peng RD. Reproducible research in computational science. *Science*. 2011;334:1226–1227. doi: 10.1126/science.1213847.

60. Roy Choudhury K, Gibson R. Reproducible research in medical imaging. *Mol Imaging Biol*. 2012;14:395–396. doi: 10.1007/s11307-012-0569-8.
61. LeVeque RJ, Mitchell IM, Stodden V. Reproducible research for scientific computing: tools and strategies for changing the culture. *Comput Sci Eng*. 2012;14:13–17.
62. Bosenberg M, Arnheiter H, Kelsh R. The test of time. *Pigment Cell Melanoma Res*. 2013;26:157. doi: 10.1111/pcmr.12071.
63. Ioannidis JP, Khoury MJ. Improving validation practices in "omics" research. *Science*. 2011;334:1230–1232. doi: 10.1126/science.1211811.
64. Hayden EC. Personalized cancer treatments suffer setback. *Nature* 2013; doi: 10.1038/nature.2013.14238
65. Carlin SP, Standing LG. Is intelligence enhanced by letter priming? A failure to replicate the results of Ciani and Sheldon (2010). *Psychol Rep*. 2013;112:533–544. doi: 10.2466/04.03.PR0.112.2.533-544.
66. Fanelli D. "Positive" results increase down the hierarchy of the sciences. *PLoS One*. 2010;5:e10068. doi: 10.1371/journal.pone.0010068
67. Baggerly K. More data, please! *Clin Chem*. 2013;59:459–461. doi: 10.1373/clinchem.2012.200501.
68. Poste G. Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. *Trends Mol Med*. 2012;18:717–722. doi: 10.1016/j.molmed.2012.09.003.
69. Pashler H, Wagenmakers EJ. Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspect Psychol Sci* 2012;7:528–530. doi: 10.1177/1745691612465253.
70. Bakker M, van Dijk A, Wicherts JM. The rules of the game called psychological science. *Perspect Psychol Sci* 2012;7:543–554.
71. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14:365–376. doi: 10.1038/nrn3475.
72. Nosek BA, Spies JR, Motyl, M. Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspect Psychol Sci* 2012;7:615–631.
73. Alexander A, Barnett-Cowan M, Bartmess, E, et al. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspect Psychol Sci* 2012;7:657–660.
74. Wagenmakers EJ, Forstmann BU. Rewarding high-power replication research. *Cortex*. 2014;51:105–106. doi: 10.1016/j.cortex.2013.09.010.
75. Kraus WL. Do you see what I see? Quality, reliability, and reproducibility in biomedical research. *Mol Endocrinol*. 2014;28:277–280. doi: 10.1210/me.2014-1036.
76. Hooijmans CR, Ritskes-Hoitinga M. Progress in using systematic reviews of animal studies to improve translational research. *PLoS Med*. 2013;10:e1001482. doi: 10.1371/journal.pmed.1001482.
77. Macleod MR, O'Collins T, Howells DW, Donnan GA. Pooling of animal experimental data reveals influence of study design and publication bias. *Stroke*. 2004;35:1203–1208. doi: 10.1161/01.STR.0000125719.25853.20.
78. Glaeser EL. *Researcher Incentives and Empirical Methods* [dissertation]. Harvard Institute of Economic Research. Discussion Paper Number 2122; 2006.
79. Cachat J, Bandrowski A, Grethe JS, Gupta A, Astakhov V, Imam F, Larson SD, Martone ME. A survey of the neuroscience resource landscape: perspectives from the neuroscience information framework. *Int Rev Neurobiol*. 2012;103:39–68. doi: 10.1016/B978-0-12-388408-4.00003-4.
80. Lemmon VP, Abeyruwan S, Visser U, Bixby JL. Facilitating transparency in spinal cord injury studies using data standards and ontologies. *Neural Regen Res*. 2014;9:6–7. doi: 10.4103/1673-5374.125322.
81. Landis SC, Amara SG, Asadullah K, et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012;490:187–191. doi: 10.1038/nature11556.
82. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet*. 2000;24:227–235. doi: 10.1038/73432.
83. Rae JM, Creighton CJ, Meck JM, Haddad BR, Johnson MD. MDA-MB-435 cells are derived from M14 melanoma cells–a loss for breast cancer, but a boon for melanoma research. *Breast Cancer Res Treat*. 2007;104:13–19. doi: 10.1007/s10549-006-9392-8.
84. Wadman M. NIH mulls rules for validating key results. *Nature*. 2013;500:14–16. doi: 10.1038/500014a.
85. Doshi P, Goodman SN, Ioannidis JP. Raw data from clinical trials: within reach? *Trends Pharmacol Sci*. 2013;34:645–647. doi: 10.1016/j.tips.2013.10.006.
86. Goldacre B. What the Tamiflu saga tells us about drug trials and big pharma. *The Guardian* 2014; http://www.theguardian.com/business/2014/apr/10/tamiflu-saga-drug-trials-big-pharma
87. Djulbegovic B, Hozo I, Ioannidis JP. Improving the drug development process: more not less randomized trials. *JAMA*. 2014;311:355–356. doi: 10.1001/jama.2013.283742.
88. Snow DM. Commentary on: "Facilitating transparency in spinal cord injury studies using data standards and ontologies". *Neural Regen Res*. 2014;9:8–9. doi: 10.4103/1673-5374.125323.
89. Shamoo A, Aannau Z. Ensuring scientific integrity. *Nature* 1987;327:550.
90. Van Noorden R. Meeting targets lab lapses. *Nature*. 2013;497:300–301. doi: 10.1038/497300a.
91. Echols HG. *Operators and Promoters: The Story of Molecular Biology and Its Creators*. University of California Press; 2001:341–344.
92. The Maxine Singer Papers. *Profiles in Science*, National Library of Medicine. http://profiles.nlm.nih.gov/ps/retrieve/Collection/CID/DJ
93. The Paul Berg Papers. *Profiles in Science*. National Library of Medicine. http://profiles.nlm.nih.gov/ps/retrieve/Narrative/CD/p-nid/257
94. The President. Removing barriers to responsible scientific research involving human stem cells. *Federal Register* 2009;74:10667–10668. http://www.gpo.gov/fdsys/pkg/FR-2009-03-11/pdf/E9-5441.pdf
95. Chosewood CL, Wilson DE. eds. *Biosafety in Microbiological and Biomedical Laboratories*. 5th ed. US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institutes of Health. HHS Publication Number (CDC) 21–1112; 2009 (http://www.cdc.gov/biosafety/publications/bmbl5/bmbl.pdf).
96. Shekman R. How journals like nature, cell and science are damaging science. *The Guardian*. 2013: Dec 9.
97. Announcement. Reducing our Irreproducibility. *Nature* 2013;496:398.
98. Begley CG. The test of time and editorial responsibility. *Pigment Cell Melanoma Res*. 2013;26:603–604. doi: 10.1111/pcmr.12143.
99. Ioannidis JP. Journals should publish all "null" results and should sparingly publish "positive" results. *Cancer Epidemiol Biomarkers Prev*. 2006;15:186. doi: 10.1158/1055-9965.EPI-05-0921.
100. Nicholson JM, Ioannidis JP. Research grants: conform and be funded. *Nature*. 2012;492:34–36. doi: 10.1038/492034a.
101. Ioannidis JAP. More time for research: fund people not projects *Nature* 2011;477:529–531.
102. Ioannidis JAP, Khoury MJ. Assessing value in biomedical research: the PQRST of appraisal and reward. *JAMA*. 2014;312:483–484. doi: 10.1001/jama.2014.6932.
103. Mobley A, Linder SK, Braeuer R, Ellis LM, Zwelling L. A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PLoS One*. 2013;8:e63221. doi: 10.1371/journal.pone.0063221.